

(19)



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) Numéro de publication: **0 602 717 A1**

(12)

DEMANDE DE BREVET EUROPEEN

(21) Numéro de dépôt: **93203442.4**

(51) Int. Cl.⁵: **G06F 15/80, G06K 9/66**

(22) Date de dépôt: **09.12.93**

(30) Priorité: **16.12.92 FR 9215190**

(43) Date de publication de la demande:
22.06.94 Bulletin 94/25

(84) Etats contractants désignés:
DE FR GB

(71) Demandeur: **LABORATOIRES
D'ELECTRONIQUE PHILIPS
22, Avenue Descartes
F-94450 Limeil-Brévannes(FR)**

(84) **FR**

(71) Demandeur: **PHILIPS ELECTRONICS N.V.
Groenewoudseweg 1**

NL-5621 BA Eindhoven(NL)

(84) **DE GB**

(72) Inventeur: **Gentric, Philippe, Société Civile
S.P.I.D.
156, Boulevard Haussmann
F-75008 Paris(FR)**
Inventeur: **Minot, Joel, Société Civile S.P.I.D.
156, Boulevard Haussmann
F-75008 Paris(FR)**

(74) Mandataire: **Caron, Jean et al
Société Civile S.P.I.D.
156, Boulevard Haussmann
F-75008 Paris (FR)**

(54) **Dispositif neuronal et procédé pour le construire.**

(57) En partant d'un dispositif sans neurone caché, on prélève des échantillons quelconques dans un ensemble d'échantillons d'apprentissage et on les présente en tant qu'objets à classer. A chaque fois, si la réponse n'est pas correcte, un neurone caché (H_i) est introduit avec une connexion vers le neurone de sortie (O_j) de la classe de l'échantillon, alors que si la réponse est correcte aucun neurone n'est ajouté. Pendant cette phase d'introduction des neurones cachés les neurones sont répartis en groupes en recherchant pour chacun des neurones introduits s'il tombe à l'intérieur d'un groupe existant, auquel cas il lui est incorporé, et sinon un nouveau groupe est créé autour de lui. L'appartenance à un groupe est définie en fonction de la distance au neurone "créateur".

Applications aux systèmes de reconnaissance de caractères.

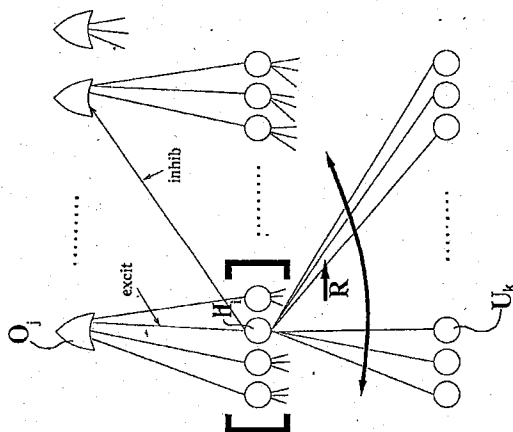


FIG. 1

EP 0 602 717 A1

La présente invention concerne un procédé pour construire un dispositif neuronal destiné au classement d'objets, au moyen d'un ensemble d'échantillons d'apprentissage ou exemples qui sont des objets dont la classe est connue, chaque objet à classer étant défini par un vecteur d'entrée représenté par un point dans un hyper-espace, le dispositif comportant une couche de neurones d'entrée correspondant
 5 chacun à une des dimensions de l'hyper-espace, une couche de neurones cachés dont les entrées sont exclusivement connectées aux neurones d'entrée et dont l'activation est basée sur les coordonnées d'un point de référence de l'hyper-espace qui est associé à ce neurone caché, et une couche de neurones de sortie correspondant chacun à une des classes, procédé consistant, en partant d'un dispositif sans neurone caché, à prélever un échantillon quelconque dans le dit ensemble d'échantillons d'apprentissage, à placer
 10 alors un premier neurone caché dans le dispositif en définissant le point de référence associé à ce neurone comme étant le point qui représente dans l'hyper-espace l'échantillon prélevé et en disposant une connexion de poids positif entre ce neurone caché et le neurone de sortie correspondant à la classe de l'échantillon, puis à prélever un nouvel échantillon dans l'ensemble et à l'appliquer au dispositif en tant qu'objet à classer et, si la réponse n'est pas correcte, à introduire dans le dispositif un nouveau neurone
 15 caché correspondant au nouvel échantillon en définissant le point de référence associé à ce neurone comme étant le point qui représente le nouvel échantillon et en disposant une connexion de poids positif entre le neurone caché et le neurone de sortie correspondant à la classe du nouvel échantillon, alors que si la réponse est correcte aucun neurone caché supplémentaire correspondant au nouvel échantillon n'est ajouté, et à traiter tous les échantillons restants de la même manière jusqu'à ce qu'il n'y ait plus
 20 d'échantillons dans l'ensemble.

Elle concerne aussi un dispositif neuronal destiné à classer des objets en J classes distinctes, les objets à classer étant définis par des vecteurs d'entrée à K coordonnées représentés par des points dans un hyper-espace à K dimensions, comportant une couche de neurones d'entrée, en nombre K, correspondant
 25 chacun à une des dites dimensions, une couche de neurones de sortie, en nombre J, correspondant chacun à une des dites classes, et une couche de neurones cachés dont l'activation est calculée en fonction des coordonnées dans l'hyper-espace d'un point de référence associé à chacun d'eux.

Un dispositif ainsi construit est utilisé notamment pour le classement d'objets, c'est-à-dire pour la reconnaissance d'objets : il s'agit par exemple de reconnaissance de formes et plus particulièrement de reconnaissance de caractères.

Un tel mode de construction est connu du document "A neural model for categories learning" de Douglas L. Reilly et autres, dans "Biological Cybernetics" n° 45, 35-41 (1982). Le réseau est construit par apprentissage incrémental et la décision pour déterminer la classe est prise sur la base d'une fonction à seuil variable qui définit dans un hyper-espace un volume de domination autour des échantillons.

L'introduction plus ou moins aléatoire des échantillons peut néanmoins avoir pour conséquence que,
 35 pour une fiabilité suffisante, le nombre de neurones cachés peut atteindre une valeur supérieure à celle qui serait obtenue pour une définition optimale de ces neurones, avec de ce fait un temps de calcul plus long lors de l'utilisation du réseau. L'invention se propose de fournir un réseau qui, construit par ce procédé, ait une très grande rapidité lors de son utilisation.

A cet effet, selon l'invention, des groupes de neurones avec un neurone représentatif pour chaque
 40 groupe sont définis au fur et à mesure de l'introduction des neurones cachés en recherchant, pour chacun des neurones introduits, s'il fait partie d'un groupe préalablement défini auquel cas il lui est incorporé, alors que s'il ne fait partie d'aucun groupe préalablement défini un nouveau groupe est formé dont il est le neurone représentatif, le premier neurone introduit ayant donc été défini comme représentatif d'un premier groupe.

Dans ce procédé, un neurone est avantageusement considéré comme faisant partie d'un groupe donné
 45 si la distance entre le point de référence associé à ce neurone et le point de référence associé au neurone représentatif du groupe en question est inférieure à une distance donnée.

Selon un perfectionnement, les groupes en question étant considérés comme d'un premier niveau, on définit au moins une seconde distance donnée plus grande que la première, à partir de laquelle au moins
 50 des seconds groupes d'un second niveau sont définis au cours du même processus que les groupes du premier niveau.

Lorsque deux neurones cachés correspondant à deux échantillons d'apprentissage ont des points représentatifs dont les coordonnées sont voisines, mais que ces échantillons ne font néanmoins pas partie
 55 de la même classe, une connexion de poids négatif est avantageusement ajoutée entre le neurone caché correspondant à un des échantillons et le neurone de sortie correspondant à la classe de l'autre échantillon.

Ceci améliore la fiabilité des résultats de classement.

Lorsque l'activation des neurones cachés est une fonction prédéterminée, et non pas ajustable, de la distance entre les points de référence, il se peut que la sensibilité du dispositif à des différences entre

certains échantillons soit trop faible pour les distinguer correctement l'un de l'autre.

Pour résoudre ce problème, un neurone caché supplémentaire, correspondant à un échantillon d'apprentissage, est avantageusement ajouté avec une liaison vers le neurone de sortie correspondant, si pour cet échantillon présenté en tant qu'objet à classer un neurone de sortie est effectivement activé pour la classe voulue, mais que la différence entre l'activité du neurone de sortie le plus activé et celle du neurone de sortie dont l'activation est la plus forte après celle du plus activé est inférieure à un seuil prédéterminé.

Le choix des valeurs d'activation des neurones cachés et de la règle de sélection déterminant la classe en fonction de l'activation des neurones de sortie a, pour une organisation donnée du réseau, un impact important sur les performances. Dans le document cité plus haut, l'utilisation d'une fonction d'activation des neurones cachés nulle au dessous du seuil ne permet pas d'intégrer de façon progressive l'influence d'échantillons d'apprentissage "lointains", et l'utilisation de neurones "binaires" (en tout ou rien) dans la couche de sortie ne permet pas de prendre des décisions motivées. En outre on ne peut pas gérer une situation dans laquelle deux neurones de sortie sont simultanément actifs.

C'est pourquoi le procédé est plus avantageusement appliqué à un dispositif dont l'activation d'un neurone caché est une fonction décroissante, c'est-à-dire inverse de la distance géométrique dans l'hyper-espace entre le point de référence associé au neurone caché et le point représentant le vecteur d'entrée, dont la fonction d'activation d'un neurone de sortie est une fonction à seuil d'entrée, et pour lequel la classe indiquée comme résultat est celle qui correspond au neurone de sortie le plus activé à condition que la différence entre l'activité de ce dernier et celle du neurone de sortie dont l'activation est la plus forte après celle du plus activé soit supérieure à une certaine valeur.

Ceci a en outre l'avantage de pouvoir proposer non seulement un gagnant mais une liste de gagnants dans l'ordre décroissant de vraisemblance, ce qui est un avantage important en reconnaissance de caractères ou l'on peut ensuite utiliser cette liste pour corriger des fautes à partir d'un dictionnaire.

Un tel dispositif, dont les neurones cachés et leurs points de référence associés ont été déterminés et les neurones cachés ont été répartis en groupes avec dans chaque groupe un neurone représentatif du groupe par le procédé selon l'invention, est avantageusement muni de moyens pour, lors de la présentation d'un échantillon à classer, faire fonctionner dans un premier temps seulement les neurones cachés représentatifs, ceci pour tous les groupes, et déterminer lequel prend la plus forte activation, puis pour faire fonctionner, dans un deuxième temps, tous les neurones cachés du groupe ou des quelques groupes dont le(s) neurone(s) représentatif(s) a (ont) fourni la plus forte activation dans le premier temps.

Lorsque plusieurs niveaux de groupes ont été définis, le dispositif est avantageusement muni de moyens pour, lors de la présentation d'un échantillon à classer, faire fonctionner dans un premier temps seulement les neurones cachés représentatifs des groupes d'un $n^{\text{ème}}$ niveau, et déterminer lequel prend la plus forte activation, de moyens pour faire fonctionner dans un deuxième temps seulement les neurones cachés représentatifs des groupes du $(n-1)^{\text{ème}}$ niveau contenus dans le groupe du $n^{\text{ème}}$ niveau dont le neurone représentatif a fourni la plus forte activation dans le premier temps, et déterminer lequel prend la plus forte activation, et ainsi de suite jusqu'au premier niveau.

Ce dispositif est en outre remarquable en ce que, pour chacun des neurones cachés, l'activation est calculée sur la base de la distance dans l'hyper-espace entre un point de référence associé à ce neurone caché et le point représentant le vecteur d'entrée et il est muni à cet effet de moyens pour déterminer l'activation d'un neurone caché d'indice "i" par la formule

$$\text{Activation} = f\left(\text{racine } n^{\text{ième}} \left(\sum_{k=0}^{K-1} (|X_k - W_{ki}|)^n \right) / \sigma_i \right)$$

dans laquelle les W_{ki} sont les K coordonnées de l'échantillon qui a entraîné la création du neurone "i", σ_i est un coefficient ajustable propre au neurone "i" les X_k sont les K coordonnées d'un vecteur d'entrée à classer, n est un nombre entier par exemple égal à deux, et la fonction "f" est une fonction qui grandit lorsque son argument tend vers zéro, un neurone de sortie est tel qu'il soit activé seulement au delà d'un certain seuil d'activation d'un neurone caché, et des moyens sont prévus pour indiquer comme résultat la classe qui correspond au neurone de la couche de sortie le plus activé à condition que la différence entre l'activité de ce dernier et celle du neurone de sortie dont l'activation est la plus forte après celle du plus activé soit supérieure à une certaine valeur.

Avantageusement, ce dispositif est muni d'une connexion de poids négatif entre un neurone caché correspondant à un échantillon et un neurone de sortie correspondant à la classe d'un autre échantillon,

dans le cas où deux neurones cachés correspondant à deux échantillons d'apprentissage ont des points représentatifs dont les coordonnées sont voisines, mais que ces échantillons ne font pas partie de la même classe.

Ces aspects de l'invention ainsi que d'autres aspects plus détaillés apparaîtront plus clairement grâce à la description suivante d'un mode de réalisation non limitatif.

La figure 1 représente schématiquement un réseau de neurones que l'invention permet de construire.

La figure 2 illustre des points représentatifs d'échantillons dans l'hyper-espace, deux dimensions étant représentées parmi les K dimensions.

Le dispositif neuronal dont le schéma est représenté sur la figure 1 est destiné à réaliser des classements. On entend par là que, en présence d'une population d'objets qui peuvent être répartis en classes distinctes, lorsqu'on présente la description d'un échantillon de cette population au dispositif, ce dernier détermine à quelle classe appartient l'échantillon. Par exemple, les objets peuvent être des lettres de l'alphabet présentées sous différentes formes ou apparences, et les classes sont : "lettre a", "lettre b", etc.

Une donnée d'entrée, qu'il s'agisse d'un échantillon d'apprentissage dont la classe est connue, ou d'un échantillon dont la classe doit être déterminée par le dispositif, se présente sous la forme d'un vecteur à K coordonnées ou composantes, dans un hyper-espace à K dimensions. Dans le cas de lettres par exemple, on peut définir une zone dans laquelle la lettre est censée se trouver et analyser cette zone par points successifs, la luminance de chacun de ces points fournissant une des composantes du vecteur d'entrée.

Le dispositif comporte une couche de neurones d'entrée U_k , en nombre égal au nombre K des coordonnées d'un vecteur d'entrée, à chacun desquels est amenée une valeur d'entrée correspondant à l'une des composantes d'un vecteur d'entrée.

Chacun des neurones H_i d'une couche de neurones cachés a une entrée reliée à la sortie d'un des neurones d'entrée U_k . Pour ne pas encombrer la figure, seules sont représentées les liaisons des neurones U vers un des neurones H_i , mais en réalité tous les neurones U sont reliés à tous les neurones H. Par contre aucune sortie d'un neurone caché n'est reliée à une entrée d'un autre neurone caché.

Le nombre N_i des neurones cachés est fonction de la difficulté du problème de classement à résoudre, comme cela sera expliqué plus loin.

Le dispositif comporte enfin une couche de neurones de sortie O_j , en nombre J, égal au nombre des classes. Plus loin sera expliqué quels neurones cachés H sont reliés à quels neurones de sortie O.

L'activation d'un neurone caché H_i est calculée sur la base de la distance, dans l'hyper-espace, entre le point représentant le vecteur d'entrée appliqué aux neurones U_k et un point de référence qui est associé au neurone H_i .

De tels points sont représentés sur la figure 2. Il s'agit d'une projection de l'espace sur le plan de la figure. Les points représentent chacun un échantillon d'apprentissage.

A chacun des points est associé un domaine représenté par un petit cercle en trait plein, entourant le point. Ce cercle est le lieu des points qui sont à une distance donnée de l'échantillon. Cette distance est représentée par la formule suivante :

$$\text{Distance (X,W)} = \text{racine } n^{\text{ième}} \left(\sum_{k=0}^{K-1} (|X_k - W_{ki}|) \right)^n$$

où les X_k sont les coordonnées d'un point de l'espace dont on cherche la distance et les W_{ki} sont les K coordonnées du point correspondant au neurone H_i , par rapport auquel on cherche la distance (X,W). Le nombre n, dans une métrique Euclidienne, est égal à 2 (dans ce cas il est inutile de prendre la valeur absolue de la différence $X_k - W_{ki}$). Toute métrique peut être employée, pourvu que dans cette métrique, on ait :

Distance (x,y) = distance (y,x)

Distance (x,z) <= distance (x,y) + distance (y,z)

La condition : distance (x,x) = 0 n'est pas nécessaire.

Pour construire un tel dispositif, il est connu de répartir les échantillons en amas ("clustering") et de mettre en place un seul neurone caché représentant chaque amas. Tout le problème consiste alors à déterminer comment grouper les échantillons en amas pour ne pas trop perdre d'information et avoir la plus grande fiabilité possible pour les résultats de classification. Avec le procédé de la présente invention,

l'élimination de ceux des échantillons qui n'ont pas besoin d'être pris en compte se fait automatiquement et aucune information n'est perdue.

On part d'un réseau qui ne comporte aucun neurone caché. On prélève d'abord n'importe quel échantillon dans l'ensemble des échantillons d'apprentissage. Cet échantillon est donc retiré de l'ensemble : on ne le rencontrera plus à l'avenir. On place un premier neurone caché correspondant à cet échantillon et l'on définit l'activation de ce neurone en fonction de la position, dans l'hyperespace, du point représentant l'échantillon, d'après la formule :

$$\text{Activation} = f(\text{racine } n^{\text{ième}} (\sum_{k=0}^{K-1} (|X_k - W_{k1}|)^n)) / \sigma_1$$

où les X_k sont les coordonnées d'un vecteur d'entrée à classer, la fonction f est une fonction qui grandit lorsque son argument tend vers zéro, par exemple une fonction de type " $\exp(-x)$ " ou une fonction de type " $1/(1+x+a)$ ", les W_{k1} sont les coordonnées du point 1 représentant le premier échantillon, et σ_1 est un coefficient ajustable propre au premier neurone.

On relie ce neurone caché H_1 au neurone de sortie O_1 correspondant à la classe de l'échantillon.

Ensuite on prend un second échantillon (par exemple au hasard) dans l'ensemble et on le retire de l'ensemble (pour ne plus le retrouver par la suite). On l'applique au dispositif en tant qu'objet à classer, et le dispositif indique donc une classe dans laquelle il range l'échantillon. Deux cas sont possibles : soit cette classe est effectivement celle de ce second échantillon, soit elle ne l'est pas. Si la classe est la bonne, on n'installe pas de neurone, et on poursuit le processus en prenant un troisième échantillon. Dans le cas où la classe indiquée n'est pas correcte, alors on place un second neurone caché dans le dispositif dont l'activation est définie par :

$$\text{Activation} = f(\text{racine } n^{\text{ième}} (\sum_{k=0}^{K-1} (|X_k - W_{k2}|)^n)) / \sigma_2$$

les W_{k2} étant les coordonnées du point représentant le second échantillon donnant lieu à la création d'un neurone, et σ_2 est un coefficient ajustable propre au second neurone, et on relie ce second neurone caché au neurone de sortie correspondant à la classe du second échantillon.

On continue ainsi en prenant tour à tour dans un ordre quelconque tous les échantillons restants, en les appliquant au dispositif en tant qu'objets à classer, avec à chaque fois création d'un nouveau neurone seulement si la classe indiquée n'est pas correcte, ce neurone ayant l'activation :

$$\text{Activation} = f(\text{racine } n^{\text{ième}} (\sum_{k=0}^{K-1} (|X_k - W_{ki}|)^n)) / \sigma_i$$

où les W_{ki} sont les coordonnées de l'échantillon qui a entraîné la création du neurone i , σ_i étant un coefficient ajustable propre au neurone i .

On continue ainsi jusqu'à ce qu'il n'y ait plus d'échantillons dans l'ensemble. Le dispositif se construit donc de lui-même en mettant en place les neurones cachés et leurs liaisons avec les neurones de sortie. Plus la classe des exemples est difficile à déterminer, c'est-à-dire plus le problème est difficile, plus il y a d'exemples pour lesquels une bonne réponse n'est pas trouvée d'emblée (lorsque les exemples sont appliqués au dispositif en tant qu'objets à classer), et plus le nombre de neurones cachés introduits est grand.

Dans le cas où le réseau de neurones n'est pas réel, mais est simulé sur un ordinateur conventionnel, les calculs déterminant la classe d'un échantillon présenté comme objet à classer sont particulièrement rapides au début du processus puisqu'il y a très peu de neurones. Le processus ralentit ensuite au fur et à mesure de l'introduction de nouveaux neurones, mais il reste néanmoins globalement très rapide.

Considérons les points D et H voisins de la figure 2, et supposons qu'ils font partie de classes différentes. Il est possible que, lors de leurs examens comme objets à classer, les résultats obtenus, du fait du voisinage, soient les mêmes pour ces deux points. Bien entendu, ce cas donne lieu à la création d'un

neurone, au moins pour l'échantillon examiné en second, mais il est en outre avantageux de compléter le processus en ajoutant dans ce cas, en plus de la synapse normale établie entre le neurone caché ainsi créé et le neurone de sortie de la classe trouvée par erreur, une synapse de poids négatif. Avec cette façon de faire, le nombre total de synapses de sortie est alors au plus égal à deux fois le nombre de neurones cachés.

Avec le réseau tel que décrit jusqu'ici, lorsqu'un objet est présenté pour être classé, toutes les distances (dans l'hyper-espace) entre le point correspondant à l'objet présenté et les points correspondant à tous les neurones cachés doivent être déterminées. En simulation sur un ordinateur conventionnel, ces opérations sont coûteuses. Pour réduire le temps de calcul, il est proposé de répartir les échantillons en groupes et dans un premier temps d'utiliser seulement les neurones cachés représentatifs d'un groupe, puis dans un deuxième temps de rechercher une activation suffisante, ou bien la plus forte activation, parmi les neurones cachés du groupe dont le neurone représentatif a fourni la plus forte activation dans le premier temps. Ce processus en deux temps sera expliqué de façon plus détaillée plus loin.

Dans l'exemple de la figure 2, les groupes sont délimités par des hyper-sphères A, B, C, qui sont définies au fur et à mesure de l'introduction des neurones cachés. Pour cela, au cours du processus d'introduction des neurones cachés, on recherche pour chacun des neurones introduits s'il fait partie d'un groupe préalablement défini auquel cas il lui est incorporé, alors que s'il ne fait partie d'aucun groupe préalablement défini un nouveau groupe est formé dont il est le neurone représentatif. Le premier neurone introduit est par conséquent défini comme représentatif du premier groupe. Un neurone est considéré comme faisant partie d'un groupe donné si la distance entre le point de référence associé à ce neurone et le point de référence associé au neurone représentatif du groupe en question est inférieure au rayon d'une hyper-sphère de rayon déterminé centrée sur le point de référence correspondant au neurone représentatif du groupe, autrement dit s'il est dans cette hyper-sphère. Le rayon en question est une grandeur de même nature que le paramètre σ_1 défini plus haut. Ainsi par exemple on peut imaginer que, lors de la construction du réseau, le point F est rencontré le premier, aucun des autres points de l'hyper-sphère A n'ayant été trouvé jusque là. Une hyper-sphère A est alors définie, centrée sur le point F. Lorsque le point E est rencontré, il ne fait pas partie de l'hyper-sphère A (ni d'aucune autre), donc une hyper-sphère B est créée. Lorsqu'ensuite le point G est rencontré, comme il ne fait pas partie des hyper-sphères précédentes, l'hyper-sphère C est créée. Les autres points des hyper-sphères ont été rencontrés après le point F pour ce qui concerne les points de l'hyper-sphère A, après le point E pour ce qui concerne les points de l'hyper-sphère B, et après le point G pour ce qui concerne les points de l'hyper-sphère C. Il faut remarquer que si le point D par exemple avait été rencontré avant le point G, une hypersphère aurait été créée avec le point D comme centre. Le point G y serait contenu, et de ce fait l'hyper-sphère C n'aurait pas été créée.

Ainsi donc au cours du processus d'introduction des neurones cachés, on regarde pour chaque exemple non seulement quelle est la réponse (le classement) fournie par le dispositif, mais encore, au cas où cet exemple conduit à l'introduction d'un nouveau neurone, si le point correspondant est déjà dans une hyper-sphère existante. La répartition en groupes est ainsi déjà établie à l'issue du processus principal sans que l'on ait à procéder à une passe supplémentaire.

On constate que la façon dont sont définis les groupes dépend de l'ordre dans lequel sont rencontrés les échantillons d'apprentissage, le hasard intervenant donc dans le processus. Des hyper-sphères peuvent ainsi être amenées à se recouper. Par exemple, le point H est défini indifféremment comme faisant partie du groupe B ou du groupe C selon que, lors du calcul des distances du point H aux centres des hypersphères, le point E ou le point G a été rencontré le premier. On a constaté que ceci n'avait pas d'inconvénient et ne diminuait pas les performances du dispositif final, notamment grâce au mode d'utilisation de ce dernier.

On a décrit ci-dessus une manière de définir un certain niveau de groupement des neurones cachés. On entend par "niveau de groupement" le fait qu'un groupe comprenne un plus ou moins grand nombre de neurones, autrement dit que le rayon de l'hyper-sphère définissant le groupe est plus ou moins grand. Il est clair qu'un deuxième niveau ou plusieurs autres niveaux pourraient être déterminés de la même façon sans difficultés. Il suffirait à cet effet de définir un ou plusieurs autres rayons (de plus en plus grands) pour des hyper-sphères centrées sur le point de référence correspondant au neurone représentatif d'un groupe (le critère d'appartenance à un groupe étant le fait pour un point d'être dans cette hyper-sphère), et à partir desquels des groupes de plusieurs niveaux pourraient être définis au cours du même processus.

Lorsque l'on utilise le dispositif, la classe d'un échantillon est déterminée de la manière suivante:

- dans un premier temps on utilise seulement les neurones cachés représentatifs d'un groupe, par exemple tous les neurones correspondant aux points situés dans la même hyper-sphère A sont représentés par le neurone caché F, ceux situés dans la même hyper-sphère B sont représentés par le neurone caché E, et ceux situés dans la même hypersphère C sont représentés par le neurone

caché G, et on détermine celui qui prend la plus forte activation,

- puis dans un deuxième temps on calcule l'activation des autres neurones du groupe dont le neurone représentatif a fourni la plus forte activation dans le premier temps, et dès que l'on trouve un neurone dont l'activation dépasse un seuil donné (autrement dit lorsque le point représentatif de l'échantillon est compris dans une hyper-sphère de rayon donné centrée sur le point représentatif du neurone) on considère ce neurone comme étant le gagnant. Au lieu de s'arrêter dès ce moment, on peut aussi étendre la recherche, en cherchant la plus forte activation parmi tous les neurones cachés du groupe dont le neurone représentatif a fourni la plus forte activation dans le premier temps. Si c'était par exemple le groupe C, il est possible que dans le deuxième temps ce soit le neurone correspondant au point D qui prenne la plus forte activation. Il faut bien voir qu'il est parfaitement possible que ce point D ne fasse pas partie de la même classe que le point G. L'échantillon d'entrée sera alors classé dans la classe du point D. C'est notamment pour cela que le côté aléatoire de la détermination des groupes n'a finalement pas d'inconvénient.

Il est clair que ce processus s'applique aisément au cas où il y a plus de deux niveaux de groupement des neurones cachés, en traitant les niveaux tour à tour dans l'ordre décroissant de la dimensions des groupes correspondants.

Quand on ajoute, en plus d'une synapse normale établie entre un neurone caché et un neurone de sortie, une synapse de poids négatif pour des échantillons d'apprentissage qui ont des points représentatifs voisins, mais qui ne font pas partie de la même classe, la règle pour déterminer si ces points sont voisins peut par exemple être que les deux exemples de classes différentes font partie du même groupe.

On comprend aisément que le temps de calcul est réduit, puisque si par exemple il y a 900 neurones cachés divisés en 30 groupes de 30 neurones, on aura à calculer 30 + 30 activations au lieu de 900.

On peut en outre étendre la recherche en recherchant la plus forte activation non pas parmi les neurones cachés d'un seul groupe, mais parmi ceux des quelques groupes dont les neurones représentatifs ont fourni les plus fortes activations dans le premier temps. Cela allonge un peu le calcul mais procure plus de sécurité. On peut par exemple choisir arbitrairement un nombre prédéterminé de groupes dont on considérera tous les neurones au cours du deuxième temps, ou bien fixer un seuil d'activation minimal au cours du premier temps pour être "admis" lors du deuxième temps.

Un autre moyen d'étendre la recherche est de "forcer" un rayon plus grand que le rayon choisi au départ, pour l'hyper-sphère définissant un groupe d'un niveau déterminé.

Lorsque l'on relie un neurone caché au neurone de sortie adéquat, on installe arbitrairement un poids positif qui est donc le même pour toutes ces liaisons. Les poids en question, ainsi que les coefficients ajustables σ_i , peuvent être affinés ultérieurement, lorsque tous les neurones cachés sont installés, par exemple au moyen d'un processus connu de rétro-propagation de l'erreur, qui ne demande pas un temps d'exécution excessif, du fait que l'on dispose déjà d'une solution approchée. De plus on peut utiliser la répartition en groupes pour accélérer le processus de rétro-propagation de l'erreur. Dans un tel processus, on présente un échantillon à l'entrée, on évalue le réseau (propagation), on calcule l'erreur, puis on la "rétro-propage" en modifiant les poids au fur et à mesure. Avec la répartition en groupes, on gagne du temps à l'évaluation puisqu'on n'évalue qu'une partie du réseau, mais on en gagne aussi lors de la rétro-propagation puisqu'on ne modifie pas tous les poids et les σ_i du réseau mais seulement ceux qui concernent les sorties reliant les neurones cachés du ou des groupe(s) utile(s), dont on avait mémorisé la liste au cours de la propagation.

Pour limiter les inconvénients éventuels dûs au fait que les neurones cachés ont été créés un peu au hasard il est en outre possible, une fois le dispositif terminé, de mettre en oeuvre un processus "d'élagage", aux fins d'enlever des neurones superflus. Ceci peut se faire par exemple en proposant à nouveau les exemples d'apprentissage comme échantillons à reconnaître, mais en désactivant à chaque fois le neurone caché qui avait été mis en place à propos de cet exemple. Si le résultat est néanmoins correct, on supprime alors ce neurone.

50 Revendications

1. Procédé pour construire un dispositif neuronal destiné au classement d'objets, au moyen d'un ensemble d'échantillons d'apprentissage ou exemples qui sont des objets dont la classe est connue, chaque objet à classer étant défini par un vecteur d'entrée représenté par un point dans un hyper-espace, le dispositif comportant une couche de neurones d'entrée correspondant chacun à une des dimensions de l'hyper-espace, une couche de neurones cachés dont les entrées sont exclusivement connectées aux neurones d'entrée et dont l'activation est basée sur les coordonnées d'un point de référence de l'hyper-espace qui est associé à ce neurone caché, et une couche de neurones de sortie

- correspondant chacun à une des classes, procédé consistant, en partant d'un dispositif sans neurone caché, à prélever un échantillon quelconque dans le dit ensemble d'échantillons d'apprentissage, à placer alors un premier neurone caché dans le dispositif en définissant le point de référence associé à ce neurone comme étant le point qui représente dans l'hyper-espace l'échantillon prélevé et en disposant une connexion de poids positif entre ce neurone caché et le neurone de sortie correspondant à la classe de l'échantillon, puis à prélever un nouvel échantillon dans l'ensemble et à l'appliquer au dispositif en tant qu'objet à classer et, si la réponse n'est pas correcte, à introduire dans le dispositif un nouveau neurone caché correspondant au nouvel échantillon en définissant le point de référence associé à ce neurone comme étant le point qui représente le nouvel échantillon et en disposant une connexion de poids positif entre le neurone caché et le neurone de sortie correspondant à la classe du nouvel échantillon, alors que si la réponse est correcte aucun neurone caché supplémentaire correspondant au nouvel échantillon n'est ajouté, et à traiter tous les échantillons restants de la même manière jusqu'à ce qu'il n'y ait plus d'échantillons dans l'ensemble, caractérisé en ce que des groupes de neurones avec un neurone représentatif pour chaque groupe sont définis au fur et à mesure de l'introduction des neurones cachés en recherchant, pour chacun des neurones introduits, s'il fait partie d'un groupe préalablement défini auquel cas il lui est incorporé, alors que s'il ne fait partie d'aucun groupe préalablement défini un nouveau groupe est formé dont il est le neurone représentatif, le premier neurone introduit ayant donc été défini comme représentatif d'un premier groupe.
2. Procédé selon la revendication 1, caractérisé en ce qu'un neurone est considéré comme faisant partie d'un groupe donné si la distance entre le point de référence associé à ce neurone et le point de référence associé au neurone représentatif du groupe en question est inférieure à une distance donnée.
3. Procédé selon la revendication 2 caractérisé en ce que, les groupes en question étant considérés comme d'un premier niveau, on définit au moins une seconde distance donnée plus grande que la première, à partir de laquelle au moins des seconds groupes d'un second niveau sont définis au cours du même processus que les groupes du premier niveau.
4. Procédé selon l'une quelconque des revendications 1 à 3, caractérisé en ce que, lorsque deux neurones cachés correspondant à deux échantillons d'apprentissage ont des points représentatifs dont les coordonnées sont voisines, mais que ces échantillons ne font néanmoins pas partie de la même classe, une connexion de poids négatif est ajoutée entre le neurone caché correspondant à un des échantillons et le neurone de sortie correspondant à la classe de l'autre échantillon.
5. Procédé selon l'une quelconque des revendications 1 à 4, caractérisé en ce qu'un neurone caché supplémentaire, correspondant à un échantillon d'apprentissage, est ajouté avec une liaison vers le neurone de sortie correspondant, si pour cet échantillon présenté en tant qu'objet à classer un neurone de sortie est effectivement activé pour la classe voulue, mais que la différence entre l'activité du neurone de sortie le plus activé et celle du neurone de sortie dont l'activation est la plus forte après celle du plus activé est inférieure à un seuil prédéterminé.
6. Dispositif neuronal destiné à classer des objets en J classes distinctes, les objets à classer étant définis par des vecteurs d'entrée à K coordonnées représentés par des points dans un hyper-espace à K dimensions, comportant une couche de neurones d'entrée, en nombre K, correspondant chacun à une des dites dimensions, une couche de neurones de sortie, en nombre J, correspondant chacun à une des dites classes, et une couche de neurones cachés dont l'activation est calculée en fonction des coordonnées dans l'hyper-espace d'un point de référence associé à chacun d'eux, caractérisé en ce que les neurones cachés et leurs points de référence associés ayant été déterminés et les neurones cachés ayant été répartis en groupes avec dans chaque groupe un neurone représentatif du groupe par le procédé selon la revendication 1, l'activation d'un neurone caché étant une fonction inverse de la distance géométrique dans l'hyper-espace entre le point de référence associé au neurone caché et le point représentant le vecteur d'entrée, la fonction d'activation d'un neurone de sortie étant une fonction à seuil d'entrée, et la classe indiquée comme résultat étant celle qui correspond au neurone de sortie le plus activé à condition que la différence entre l'activité de ce dernier et celle du neurone de sortie dont l'activation est la plus forte après celle du plus activé soit supérieure à une certaine valeur, il est muni de moyens pour, lors de la présentation d'un échantillon à classer, faire fonctionner dans un premier temps seulement les neurones cachés représentatifs des groupes, et déterminer lequel prend la plus forte activation, puis pour faire fonctionner, dans un deuxième temps, tous les neurones cachés

du groupe ou des quelques groupes dont le(s) neurone(s) représentatif(s) a (ont) fourni la plus forte activation dans le premier temps.

- 5 7. Dispositif selon la revendication 6, caractérisé en ce que, les neurones cachés ayant été en outre répartis en plusieurs groupes par le procédé selon la revendication 3, il est muni de moyens pour, lors de la présentation d'un échantillon à classer, faire fonctionner dans un premier temps seulement les neurones cachés représentatifs des groupes d'un $n^{\text{ième}}$ niveau, et déterminer lequel prend la plus forte activation, de moyens pour faire fonctionner dans un deuxième temps seulement les neurones cachés représentatifs des groupes du $(n-1)^{\text{ième}}$ niveau contenus dans le groupe du $n^{\text{ième}}$ niveau dont le neurone représentatif a fourni la plus forte activation dans le premier temps, et déterminer lequel prend la plus forte activation, et ainsi de suite jusqu'au premier niveau.
- 10 8. Dispositif neuronal selon l'une des revendications 6 ou 7, caractérisé en ce que, pour chacun des neurones cachés, l'activation est calculée sur la base de la distance dans l'hyper-espace entre un point de référence associé à ce neurone caché et le point représentant le vecteur d'entrée et il est muni à cet effet de moyens pour déterminer l'activation d'un neurone caché d'indice "i" par la formule :
- 15

$$20 \quad \text{Activation} = f\left(\text{racine } n^{\text{ième}} \left(\sum_{k=0}^{K-1} (|X_k - W_{ki}|)^n \right) / \sigma_i \right)$$

dans laquelle les W_{ki} sont les K coordonnées de l'échantillon qui a entraîné la création du neurone "i", σ_i est un coefficient ajustable propre au neurone "i", les X_k sont les K coordonnées d'un vecteur d'entrée à classer, n est un nombre entier, et la fonction "f" est une fonction qui grandit lorsque son argument tend vers zéro, un neurone de sortie est tel qu'il soit activé seulement au delà d'un certain seuil d'activation d'un neurone caché, et des moyens sont prévus pour indiquer comme résultat la classe qui correspond au neurone de la couche de sortie le plus activé à condition que la différence entre l'activité de ce dernier et celle du neurone de sortie dont l'activation est la plus forte après celle du plus activé soit supérieure à une certaine valeur.

25 30

9. Dispositif selon la revendication précédente, caractérisé en ce que le nombre entier n est égal à deux.
- 35 10. Dispositif selon l'une quelconque des revendications 6 à 10, caractérisé en ce qu'il est muni d'une connexion de poids négatif entre un neurone caché correspondant à un échantillon et un neurone de sortie correspondant à la classe d'un autre échantillon, dans le cas où deux neurones cachés correspondant à deux échantillons d'apprentissage ont des points représentatifs dont les coordonnées sont voisines, mais que ces échantillons ne font pas partie de la même classe.
- 40

45

50

55

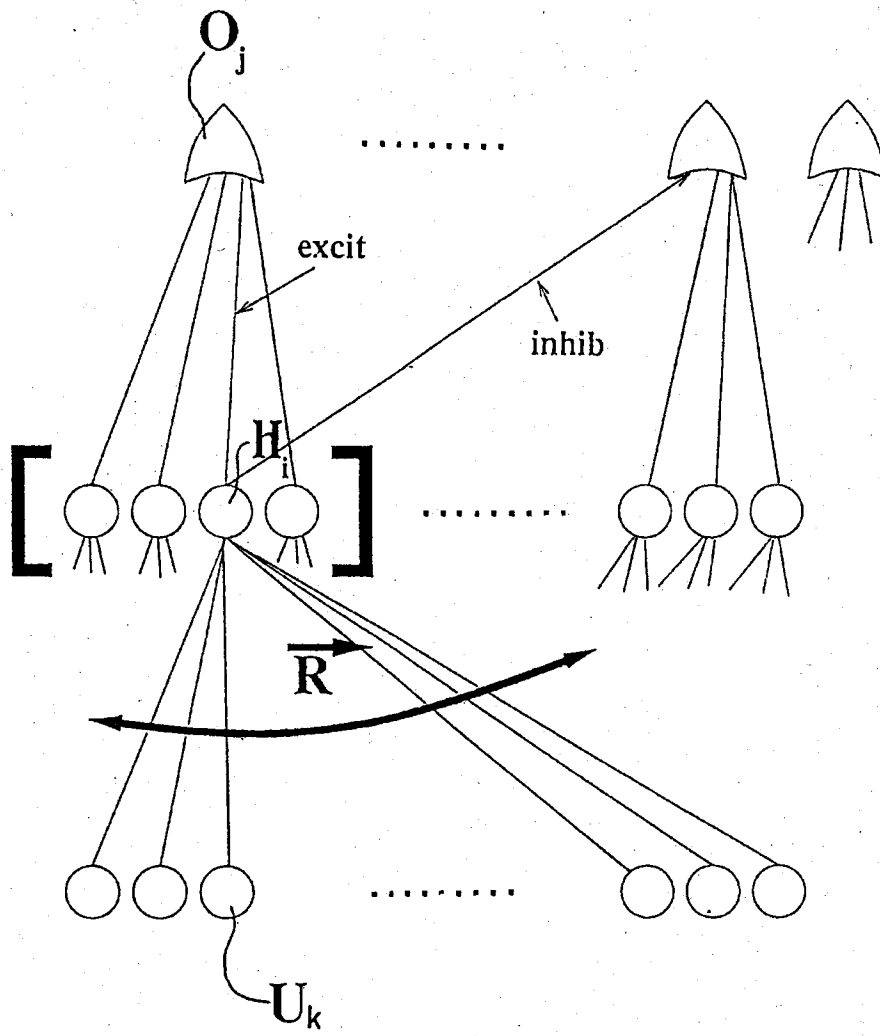


FIG. 1

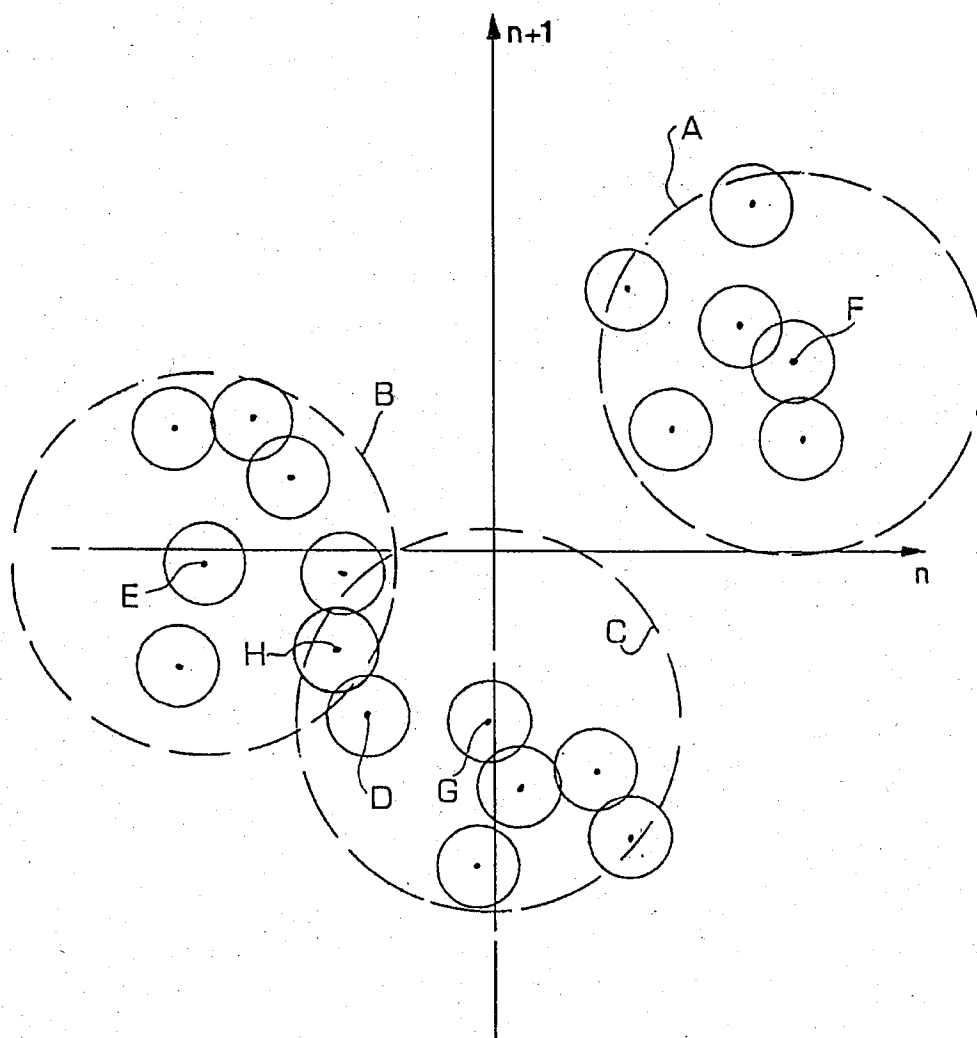


FIG. 2



Office européen
des brevets

RAPPORT DE RECHERCHE EUROPEENNE

Numero de la demande
EP 93 20 3442

DOCUMENTS CONSIDERES COMME PERTINENTS			
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	Revendication concernée	CLASSEMENT DE LA DEMANDE (Int.Cl.5)
Y	IJCNN INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS vol. 3 , 17 Juin 1990 , SAN DIEGO , USA pages 245 - 250 KONG 'The piecewise linear neural network : training and recognition' * abrégé *	1	G06F15/80 G06K9/66
A	---	2-10	
Y	1991 IEEE INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS vol. 3 , 18 Novembre 1991 , SINGAPORE pages 2709 - 2715 FANG 'Unsupervised learning for neural trees'	1	
A	* page 2709, ligne 1 - page 2714, ligne 4; figures 1,2 *	2-10	
A	---	1	
	IJCNN INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS vol. 1 , 17 Juin 1990 , SAN DIEGO , USA pages 423 - 428 CASASENT 'Adaptive clustering neural net for piecewise nonlinear discriminant surfaces' * page 423, ligne 1 - page 426, ligne 20; figures 1-4 *		DOMAINES TECHNIQUES RECHERCHES (Int.Cl.5) G06F G06K
A	---	1	
	IJCNN INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS vol. 1 , 7 Juin 1992 , BALTIMORE , USA pages 576 - 581 WENG 'Cresceptron : A self-organizing neural network which grows adaptively' * abrégé * * page 579, ligne 21 - page 581, ligne 8 *		

	-/--		
Le présent rapport a été établi pour toutes les revendications			
Lien de la recherche LA HAYE		Date d'achèvement de la recherche 1 Février 1994	Examinateur Schenkels, P
CATEGORIE DES DOCUMENTS CITES X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire T : théorie ou principe à la base de l'invention E : document de brevet antérieur, mais publié à la date de dépôt ou après cette date D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant			



Office européen
des brevets

RAPPORT DE RECHERCHE EUROPEENNE

Numero de la demande
EP 93 20 3442

DOCUMENTS CONSIDERES COMME PERTINENTS			
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	Revendication concernée	CLASSEMENT DE LA DEMANDE (Int.Cl.5)
A	IJCNN INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS vol. 1, 17 Juin 1990, SAN DIEGO, USA pages 637 - 642 DEFFUANT 'Neural units recruitment algorithm for generation of decision trees' * page 637, ligne 1 - ligne 33 * -----	1	
			DOMAINES TECHNIQUES RECHERCHES (Int.Cl.5)
Le présent rapport a été établi pour toutes les revendications			
Lieu de la recherche LA HAYE		Date d'achèvement de la recherche 1 Février 1994	Examineur Schenkels, P
CATEGORIE DES DOCUMENTS CITES			
X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire		T : théorie ou principe à la base de l'invention E : document de brevet antérieur, mais publié à la date de dépôt ou après cette date D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant	